Plant Breeding **125**, 493—500 (2006)

Characterization of the *MIPS* gene family in *Glycine max*

A. S. Chappell[1], A. M. Scaboo[2], X. Wu[3], H. Nguyen[3], V. R. Pantalone[2] and K. D. Bilyeu[1,4]

[1]USDA-ARS, Plant Genetics Research Unit, 110 Waters Hall, Columbia, MO 65211, USA; [2]Department of Plant Sciences, 2431 Joe Johnson Drive, University of Tennessee, Knoxville, TN 37996-4561, USA; [3]Division of Plant Sciences, National Center for Soybean Biotechnology, 1–31 Agriculture Building, University of Missouri, Columbia, MO 65211, USA; [4]Corresponding author, E-mail: bilyeuk@missouri.edu

With 3 figures and 4 tables

Received February 28, 2006/Accepted April 7, 2006
Communicated by W. E. Weber

Abstract

Phytic acid (*myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate) is the primary storage component of phosphorus in plant seeds. The first step in phytic acid biosynthesis is the *de novo* synthesis of *myo*-inositol, which is catalyzed by the enzyme D-*myo*-inositol 3-phosphate synthase (MIPS EC 5.5.1.4). Previous work detected four *MIPS* genes in soybean (*Glycine max*). However, only a limited amount of data were available for the *MIPS* gene family and some of the data were conflicting. The work described here clears up these data and characterizes the *MIPS* gene family for the purposes of reverse genetic technologies. The complete genomic sequence of all four genes was determined and their expression profile was examined by quantitative real-time reverse transcription-polymerase chain reaction. Our results indicate that the four *MIPS* genes are highly conserved and temporally and spatially expressed. The *MIPS* gene family in the low phytic acid soybean line, CX1834, was also characterized since this line displays a phenotype similar to previously characterized *MIPS* mutants. These data demonstrate that mutations in *MIPS* genes are not the cause of the low phytic acid phenotype.

Key words: CX1834 — *Glycine max* — MIPS — phytate — phytic acid — soybean

Myo-inositol plays critical and diverse biological roles in a myriad of cellular processes including signal transduction, stress response, cell-wall biogenesis, growth regulation, osmotolerance, IAA metabolism and phosphatidylinositol and inositol phosphate biosynthesis (for review see Loewus and Murthy 2000, Irvine and Schell 2001). The sole synthetic source of *myo*-inositol is through the conversion of glucose 6-phosphate to D-*myo*-inositol 3-phosphate, the precursor of all remaining inositol-containing compounds in the cell (Loewus and Loewus 1983). The only enzyme that converts glucose 6-phosphate to D-*myo*-inositol 3-phosphate is the isomerase, D-*myo*-inositol 3-phosphate synthase (MIPS). MIPS is a highly conserved enzyme that has been identified throughout all biological kingdoms in such diverse organisms as Archaebacteria, Eubacteria, yeast, plants, flies and humans (Majumder et al. 1997, 2003, Bachhawat and Mande 2000).

MIPS is of great interest from a crop science perspective given its key role in phytic acid (*myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate) biosynthesis. Phytic acid, a *myo*-inositol derivative, is the primary storage component of phosphorus in plant seeds and can account for up to 2% of the soybean seed dry weight (Raboy et al. 1984). Phyticacid is usually deposited in protein bodies as a mixed salt (phytate), bound to mineral cations such as Fe^{3+}, Ca^{2+}, Mg^{2+}, Zn^{2+} and K^{+} (Prattley and Stanley 1982, Lott 1984). Phytic acid is degraded during seed germination by phytase enzymes thereby liberating phosphate, inositol and mineral cations for use during seedling development (Strother 1980, Raboy 1997).

Although phytic acid is an important storage molecule for growing seedlings, it poses several problems for agriculture. Monogastric livestock lack sufficient phytase activity to metabolize the phytate present in cereal seeds, leading to inadequate phosphorous uptake and the need to supplement animal feed with expensive and non-renewable inorganic phosphorus (reviewed in Brinch-Pedersen et al. 2002). The undigested phytate that is excreted in the animal's manure (which is used as fertilizer) is the leading cause of phosphorus pollution (Cromwell and Coffey 1991). Phytic acid is considered an antinutritional compound as it strongly binds essential minerals, thereby decreasing their bioavailability (Zhou and Erdman 1995). These antinutritional qualities of phytate can be further extended to human health as phytate is considered to be the most important antinutritional factor contributing to the iron deficiency suffered by over 2 billion people worldwide (Bouis 2000).

The undesirable properties of phytic acid listed above make the development and characterization of low phytate crops a high priority in agricultural research. To this end, several low phytic acid mutants have been identified in a variety of crops (reviewed in Raboy 2002). One such soybean mutant, LR33, exhibited a 50% reduction in seed phytic acid levels (Hitz et al. 2002). Further characterization of this mutant revealed a *MIPS* mutation responsible for the low phytic acid phenotype. Interestingly, the LR33 *MIPS* mutation also conferred an increase in seed sucrose levels and a decrease in seed raffinosaccharides, both highly desirable traits in soybeans. While this mutant is extremely valuable, it is not available for research or public breeding. Therefore, this work is intended to develop a similar *MIPS* mutant via TILLING that will be publicly available for research and breeding.

TILLING (Targeting Induced Local Lesions IN Genomes) is a PCR-based high-throughput molecular screening method used to identify point mutations in a gene of interest (McCallum et al. 2000b). This particular reverse genetic technology requires that the targeted gene family be fully characterized such that the entire genomic sequences of all genes are known along with their expression profiles. However, based on the literature and our initial investigations, only a limited amount of data were available for the *MIPS* gene

family and some of the data were conflicting. The objective was to clear up the conflicting data and to characterize the *MIPS* gene family for the purposes of TILLING and other reverse genetic technologies. In addition, the *MIPS* gene family was characterized in the low phytate soybean line, CX1834, since this line displays a phenotype similar to previously characterized *MIPS* mutants. CX1834 is the lower phytate soybean line of two that are available for public soybean breeding efforts (Wilcox et al. 2000). Unlike LR33, the molecular basis for the low phytate trait in CX1834 is not known. In attempt to genetically characterize CX1834, the *MIPS* gene family in CX1834 was sequenced as well as association analysis with the seed-expressed *MIPS1* gene.

Materials and Methods

Isolation and sequencing of *MIPS* genomic sequences: PCR reactions were conducted with Titanium *Taq* (BD Biosciences, Palo Alto, CA, USA) as previously described (Bilyeu et al. 2005). In some instances Titanium *Taq* was replaced with ExTaq (TaKaRa Mirus Bio, Madison, WI, USA). Unless otherwise stated, *MIPS* gene sequences were amplified from 'Williams 82' genomic DNA and PCR products were either sequenced directly or TOPO cloned with the TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA, USA) and then sequenced. Full-length *MIPS1* was amplified from 'Williams 82' genomic DNA with the primers HMIPSf and HMIPSrev. The majority of the *MIPS2* gene was amplified with the primers MIPS2f2 and aMIPSxr. The remaining 3′ end of *MIPS2* was amplified with MIPSf1 and HMIPS-rev. *MIPS3* was divided into five sections: MIPS3f/ALLMIPSR, MIPS35a/M3i4r, M3ex4if/M3i5rev, MIPS3/5f2/AMIPSx1, and Mendf1/M5endr1. Two sections of *MIPS4* were amplified from 'Williams 82' genomic DNA: MIPS3f/ALLMIPSR and MIPS3/5f2/AMIPSx1. Two different sections of *MIPS4* were amplified from BAC DNA: MIPS5f#2/ALLMIPSr2 and M5seq4/M35z1. BAC DNA was directly sequenced with MIPS35a to get the remaining section of exon 3 and BAC DNA was sequenced with Mend1 to get the very 3′ end of *MIPS4*. All primer sequences are shown in Table 1.

Discrimination of the *MIPS1a* and *MIPS1b* alleles was by PCR of genomic DNA with the primers MIPSf1 and MIPSrev, which spanned intron 9. Reaction conditions were 95°C for 10 min followed by 35 cycles of 95°C for 30 s, 62°C for 30 s, and 72°C for 30 s. *MIPS1a* alleles produced 316 bp products while *MIPS1b* alleles were 277 bp.

Sequencing of *MIPS* cDNA from CX1834-1-3/CX1834-1-6: RNA purification: RNA was made from ground tissue using TRIzol (Invitrogen) according to manufacturer's instructions with the proteoglycan and polysaccharide contamination modification. An aliquot of each RNA sample was analysed on an agarose gel to check for intact ribosomal RNA bands. RNA samples were stored at −70°C.

Sequencing of *MIPS* cDNA from CX1834-1-3/CX1834-1-6: cDNA synthesis: SuperScript III RNase H- Reverse Transcriptase (Invitrogen) was used to make cDNA from total RNA. 500 ng of oligo$(dT)_{20}$ was added to up to 5 μg RNA in 1X First Strand Buffer (Invitrogen, 50 mM Tris-HCl pH 8.3, 75 nM KCl, 3 nM $MgCl_2$) and tubes were incubated at 70°C for 10 min and then put on ice for 5 min. Prewarmed reaction buffer containing DTT (dithiothreitol, 5 mM final) and dNTPs (deoxyribonucleotide triphosphate mix, 0.5 mM final) was added and tubes were incubated at 42°C for 5 min. SuperScript III RNase H- Reverse Transcriptase (200 U) was added to each tube and incubated at 42°C for 50 min and then heat inactivated at 75°C for 15 min.

Sequencing of *MIPS* cDNA from CX1834-1-3/CX1834-1-6: *MIPS* cDNA amplification: Ten percent of the cDNA synthesis reaction was used for PCR amplification of each gene. *MIPS1*, *MIPS2* and *MIPS3* cDNA sequence was amplified from either CX1834-1-3 or CX1834-1-6 cDNA that was made from total RNA isolated from 8 mm seeds. *MIPS1* was amplified with HMIPSf/HMIPSrev. *MIPS2* cDNA sequence was amplified with MIPS2f/MIPSrev. *MIPS3* cDNA was amplified with MIPS3f/AMIPSx1. Most of the *MIPS4* cDNA sequence was isolated from CX1834-1-3 cDNA that was made from CX1834-1-3 seedlings RNA using the primers MIPS5f#2 and MIPS5r. It was not possible to get the very 5′ end of the cDNA sequence. Therefore, the genomic sequence of the 5′ end of MIPS4 was amplified with the primers MIPS5f3 and MIPS5r4 and sequenced.

Table 1: Primers used in this study for *MIPS* genomic DNA and cDNA amplification

Target	Name	Sequence (5′–3′)	Direction
MIPS1	HMIPSf	ATGTTCATCGAGAATTTTAAGGTT	Forward
MIPS1/2	HMIPSrev	ATCACTTGTACTCGAGAATCAT	Reverse
MIPS2	MIPSfl	GTTGCTACCATCCTCAGCTACC	Forward
	MIPS2f2	TGATTAATTATCAGGGTG	Forward
	aMIPSxr	TCATTTTGGTCTGACCACTCT	Reverse
	MIPS2f	ATCGTTGGGAAAAATGTT	Forward
	MIPSrev	CAGCATTGCACGCTGCTTTGA	Reverse
MIPS3	M3i4r	CATAACACACATATATAGGGT	Reverse
	M3ex4if	AAGACATTAAGTAGGTCTAAT	Forward
	M3i5rev	AGCCCTGACGAAAACACAAAGCA	Reverse
	Mendfl	CCAAGGCCCCTCTGGTAA	Forward
	M5endrl	CTCATTCAACTACAGTATCC	Reverse
	MIPS3f	AATGTTCATCGAGAGTTTC	Forward
MIPS3/4	MIPS35a	TTCTCACTTCTCAGGGGT	Forward
	MIPS3f	AATGTTCATCGAGAGTTTC	Forward
	allMIPSr	AAAGTAATTGGCTTGTTGAATCTTG	Reverse
	MIPS3/5f2	CCCTCAGAACACCTTTGTC	Forward
	AMIPSxl	AATCCAACACAAGCCCT	Reverse
MIPS4	ALLMIPSr2	GTTGCTGATATCCCATCCCC	Reverse
	M5seq4	AACCCTGATGATGTTGTGTT	Forward
	M35zl	GTAGTAGCACACGTGATAT	Reverse
	Mendl	GCTATGCTGGAAAACATCC	Forward
	MIPS5f#2	CAGGCATCAGCTATCAGAGTGGG	Forward
	MIPS5r	TCACTTGTACTCCAGGATCATGTTG	Reverse
	MIPS5f3	CCTAACGTGAAGTACACAGAGACTG	Forward
	MIPS5r4	CCCACTCTGATAGCTGATGCCTG	Reverse

Quantitative real-time RT-PCR: RNA purification and RT-PCR amplification: RNA purification was conducted as described above. In addition, RNA samples were treated with DNAse I by incubating 2 U of DNAse I (Ambion, Austin, TX, USA) with 200 ng RNA at 37°C for 30 min and then heat inactivating at 75°C for 15 min. Quantitative real-time RT-PCR (reverse transcription-polymerase chain reaction) assays were performed in triplicate on RNA samples (with or without an RT step) using a DNA Engine Opticon2 (MJ Research, Watertown, MA, USA). Parallel amplifications were performed in separate tubes but on the same RNA pools using primers and probes to the gene of interest or the normalizer *PEPC16*. QuantiTect Probe RT-PCR Kit (Qiagen, Valencia, CA, USA) was used in 20 µl reactions containing 1X Probe RT-Master Mix, 0.5 µM final concentration of each primer, 0.1 µM final concentration of probe, 6.6 ng RNA template and 0.2 µl RT Mix. The PCR thermal cycling parameters were 50°C for 30 min, 95°C for 15 min followed by 45 cycles of 94°C for 15 s and 60°C for 1 min. RNA samples were run in the absence of a RT step to assess any genomic DNA contamination. DNA contamination was found to be at a minimum as the signal with the no RT step samples was over 1000-fold less than samples with RNA (data not shown).

Quantitative real-time RT-PCR: data analysis: Transcript levels of the *MIPS* gene family members were measured relative to the endogenous reference *PEPC16* (Tuteja et al. 2004). Calculations and statistical analysis were performed basically as described in the ABI 7700 sequence detection system User Bulletin 2 (Applied Biosystems, Foster City, CA, USA). Amplification efficiency for the four primer-probe sets was determined by amplification of genomic DNA from a dilution series using 345, 34.5, 3.45, 0.345 and 0.0345 ng or 44, 4.4, 0.44, 0.044 and 0.0044 ng.

Quantitative real-time RT-PCR: gene-specific primers and probes: Although there is extensive sequence identity among the *MIPS* gene family members, gene-specific regions were identified and used for primer design. It was verified that there was only one PCR product for each amplicon by direct sequencing of the PCR product. *MIPS* probes were ordered from Sigma-Aldrich (St Louis, MO, USA) and were labelled with the fluorescent dye FAM (6-carboxyfluorescein) and the quencher BHQ-1 (Black Hole Quencher) at the 5′ and 3′ ends, respectively. Primer sequences are shown in Table 2. *PEPC16* primer sequences have been described previously (Tuteja et al. 2004). The *PEPC16* probe was ordered from Applied Biosystems and was labelled with the fluorescent dye VIC and the quencher BHQ-1 at the 5′ and 3′ ends, respectively.

PCR-based BAC pool screening: A soybean *Bsty*I BAC library was constructed using cultivar 'Williams 82', which contains 92,160 clones, with an average insert size of 150 kb, covering 12 genome equivalents. One hundred twenty-eight 384-well microtitre plates were conceptually arranged to a cubic stack containing 32 layers with four plates per layer for the six different dimensional BAC poolings. The BAC pools consist of 208 pools, containing 49, 152 soybean BAC clones in a $48 \times 32 \times 32$ array and providing coverage of 6.6X genome equivalents. The pooling strategy followed that described by Klein et al. 2000 based on previous work by Barillot et al. 1991 and Bruno et al. 1995. Each time one of the six pool types was pooled from 128 384-well microtitre plates containing 70 µl LB media plus 12.5 µg/ml chloramphenicol per well. For BAC pooling, 40 µl of culture was removed from each well using 12-channel multi-pipettor and Matrix 8-channel expandable pipettor (Matrix Tech Corp., Hudson, NH, USA). BAC DNA isolation was performed using QIAGEN Large-Construct Kit (QIAGEN Inc.). Gene-specific primers were used to perform PCR against the 208 pools plus 'Williams 82' genomic DNA as controls. The PCR products were resolved on 3% Super Fine Resolution (SFR) agarose (Amresco, Solon, OH, USA), 1× TBE gels containing 2 µg/ml ethidium bromide. SSR gel images were scanned using an AlphaImager 2200 (Alpha Innotech Corp., San Leandro, CA, USA). The Resolve Script software developed in the sorghum physical mapping project (Klein et al. 2000) was used to de-convolute the data, with some modifications to accommodate our pool layout.

Association test for CX1834-1-2 *MIPS1a* allele: CX1834-1-2, CX1834-1-3, and CX1834-1-6 are $F_{3:5}$ low phytate lines which were developed by the USDA and Purdue University from a cross of 'Athow' × M153-1-4-6-14 (Wilcox et al. 2000). A set of 187 F_5-derived-random recombinant inbred lines (RILs) from the cross '5601T' × CX1834-1-2 was developed at the University of Tennessee using single-seed descent. '5601T' is a University of Tennessee F_6-derived MG V cultivar derived from the cross of 'Hutcheson' × TN89-39 (Pantalone et al. 2003). DNA was isolated from a random $F_{5:7}$ plant, and QTL markers for LG N and L (Satt 237 and Satt 561) were assayed as described (Walker et al. 2006). Quantitation of available phosphate levels for each line was as previously described (Walker et al. 2006). A subset of the lines was used to assay for the presence of *MIPS1a* or *MIPS1b* alleles.

Results

Genomic characterization of the *Glycine max MIPS* gene family

To obtain the complete genomic sequence of all soybean *MIPS* genes, at first the literature and GenBank were consulted for any previous characterization of the *MIPS* gene family in soybean. However, based on the literature, GenBank and the initial investigations of this work, only a limited amount of data was available for the *MIPS* gene family and some of the data was conflicting. For example, four *MIPS* genes have been detected in soybean by Southern blot, but only one gene, *MIPS1*, has been sequenced (Hegeman et al. 2001, Hitz et al. 2002). Although the complete genomic sequence of *MIPS1* was determined by Hegeman et al. 2001, only the coding sequence was made available. Upon further review, it was discovered that *MIPS1* coding sequence had been deposited in

Table 2: Primers used in this study for quantitative real-time RT-PCR

Target	Name	Sequence (5′–3′)	Direction
MIPS1	MlPSl-exp-fwd	CATCGAGAATTTTAAGGTTGAGTGTCC	Forward
	MIPSl-exp-rev	CCTAATTTAGGAACATGGATGTTGGTT	Reverse
	MIPSlprobe	[DFAM]GTCAAACCCAAATCTGTCAAATACGAA[DBH1]	Forward
MIPS2	MIPS2-exp-fwd	CATCGAGAATTTTAAGGTAGAGAGTCC	Forward
	MIPS2-exp-rev	CCCAATTTTGGAACATGGGTGTTGGTT	Reverse
	MIPS2probe	[DFAM]GTCAAACCCAAATCCGTCAACTACCAA[DBH1]	Forward
MIPS3	MIPS3-exp-fwd	CATCGAGAGTTTCAAGGTTGAGAGTCC	Forward
	MIPS3-exp-rev	CCTAGTTTAGGGACATGGGTGTTGGTT	Reverse
	MIPS3probe	[DFAM]GTCAAGCCCAAAACTGTCAAATATGAA[DBH1]	Forward
MIPS4	MIPS4-exp-fwd5	TATGTACCTTACGTTGGGGACAGCAAG	Forward
	MIPS4-exp-rev5	CTCATTTTCAGCTTTAAACTCGATTCG	Reverse
	MIPS4probe3	[DFAM]TACTCTTCTGAGATATTCATGGGCGGA[DBH1]	Forward

GenBank under two different accession numbers (AF293970 from the cultivar 'Williams 82' for Hegeman et al. 2001 and AY038802 from the cultivar 'Wye' for Hitz et al. 2002). Alignment of these two genes revealed that they differ by 6 nucleotides, representing two amino acid changes in the protein sequence. To verify which sequence was actually represented in 'Williams 82', we designed primers to the ends of the *MIPS1* coding sequence and amplified and sequenced the full-length gene from genomic DNA. The obtained exon sequence was identical to the sequence from the cultivar 'Wye' from Hitz et al. 2002.

As with *MIPS1*, the literature and GenBank provided insufficient characterization of *MIPS2*, *MIPS3* and *MIPS4* to be useful in reverse genetics applications. It was found that Hegeman et al. 2001 assembled the coding sequences of these three genes from EST data that were obtained from BLAST searching and listed the putative coding sequences in the online supplementary data. However, these data were not available online, nor were the coding sequences deposited in GenBank. Therefore, we sequenced the complete genomic sequence of *MIPS2*, *MIPS3* and *MIPS4*. As a first step towards characterizing these three genes, the set of soybean *MIPS* ESTs present in GenBank was evaluated and verified the presence of four soybean *MIPS* genes (data not shown). Primers were designed based on EST sequence and full-length genomic sequence was obtained by PCR amplification of 'Williams 82' DNA and subsequent cloning and sequencing of the PCR products. The overall genomic structure is the same among all four *MIPS* genes (Fig. 1). Furthermore, the genomic structure appears to be subject to evolutionary conservation as one of the three *MIPS* genes from *Arabidopsis* has a common exon/intron structure (At5g10170).

The complete genomic sequences of 'Williams 82' *MIPS1*, *MIPS2*, *MIPS3* and *MIPS4* have been deposited in GenBank (accession numbers DQ323904, DQ323906-DQ323908) and the protein sequences are shown in Fig. 2. Sequence analysis of these four genes demonstrated a high degree of conservation, similar to all previously identified MIPS homologues

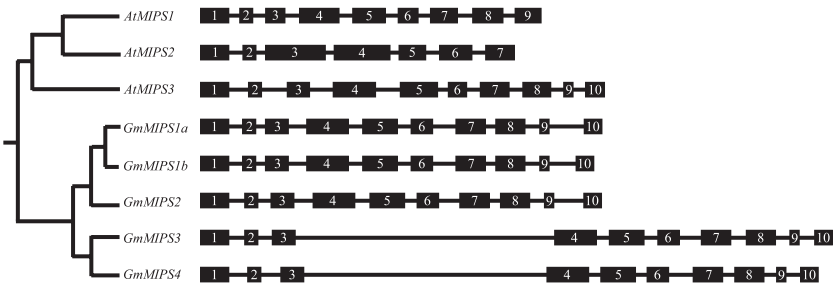

Fig. 1: Phylogenetic and genomic structure analysis of *Glycine max* and *Arabidopsis thaliana MIPS* genes. A dendrogram was produced based on a coding sequence alignment using ClustalW1.83 (http://clustalw.genome.jp/). Next to the phylogenetic tree is a representation of the genomic structure for each gene. Exons and introns are to scale and are depicted as black boxes and solid lines, respectively. Note that *GmMIPS3* and *GmMIPS4* have a significantly larger intron 3 compared with the other *MIPS* genes. *GmMIPS1a* and *GmMIPS1b* are separate alleles of the same gene. Other than *GmMIPS1b*, all soybean sequences are derived from 'Williams 82'. *AtMIPS1* is At4g39800, *AtMIPS2* is At2g22240, and *AtMIPS3* is At5g10170 (Johnson 1994, Johnson and Burk 1995, Johnson and Sussex 1995, Lackey et al. 2003)

Fig. 2: Alignment of soybean and *Arabidopsis* MIPS proteins. The open reading frames for the 'Williams 82' *MIPS* genes were translated and subjected to a CLUSTALW1.8 alignment with the *Arabidopsis* MIPS proteins. Identical amino acid residues are shaded black and similar amino acids are shaded grey

Table 3: Distribution of *MIPSI* alleles in various cultivars

PI number	Cultivar	*MIPS1* allele
88788	–	*MIPS1a*
153243	Dunfield	*MIPS1a*
157434	Illini	*MIPS1a*
189888	Mandarin (Ottawa)	*MIPS1a*
518671	Williams 82	*MIPS1a*
548298	A.K. Harrow	*MIPS1a*
548362	Lincoln	*MIPS1a*
548402	Peking	*MIPS1a*
548406	Richland	*MIPS1a*
548655	Forrest	*MIPS1a*
–	CX1834[1]	*MIPS1a*
–	M766[1]	*MIPS1a*
–	Ohio FGI[2]	*MIPS1a*
–	U95-3813[1]	*MIPS1a*
518664	Hutcheson	*MIPS1b*
548667	Essex	*MIPS1b*
630984	5601T	*MIPS1b*
–	V97-5558[1]	*MIPS1b*
–	X325[1]	*MIPS1b*

[1]Experimental breeding line with no PI number.
[2]St. Martin et al. 1996.

(Majumder et al. 1997, 2003, Bachhawat and Mande 2000). At the protein level, MIPS1 and MIPS2 are 99% identical whereas MIPS3 and MIPS4 are 96% and 95% identical to MIPS1, respectively.

A polymorphic *MIPS1* gene, *MIPS1b*, defined by a 39-bp deletion in the final intron compared with *MIPS1a* was reported to be present in some cultivars (Hitz et al. 2002). In that work, the authors tried to amplify *MIPS1b* from cDNA and reported that the coding sequence contained 40 base changes (representing seven amino acid differences) to *MIPS1a*. To validate the *MIPS1b* variant, primers were designed to amplify the genomic region surrounding the final *MIPS1* intron and DNA from different soybean cultivars were screened for the polymorphic intron containing the 39-bp deletion. The tofu breeding line V97-5558 contained the *MIPS1b* variant and was selected for further characterization of the *MIPS* gene family. The full-length *MIPS1b* gene was amplified from V97-5558 genomic DNA, cloned, and sequenced. The variant *MIPS1b* gene had only two differences with the 'Williams 82' *MIPS1a* genomic sequence: a 39 bp deletion in the final intron and a single base pair deletion in the sixth intron. No nucleotide changes within the coding sequence were detected. Sequences corresponding to the *MIPS2*, *MIPS3*, and *MIPS4* genes were confirmed to be present in the genome of V97-5558 (data not shown). These data indicate that *MIPS1b* does not contain the 40-bp changes reported by Hitz et al. 2002, at least for the allele represented in V97-5558. The complete genomic sequence of *MIPS1b* has been deposited in GenBank under the accession number DQ323905.

Eight commercial wild type soybean lines were previously examined for the presence of the *MIPS1b* allele and only two of these contained the allele (Hitz et al. 2002). In order to identify more soybean lines that contain the *MIPS1b* allele a subset of North American ancestral lines (Gizlice et al. 1994) was assayed for the presence of the *MIPS1a* or *MIPS1b* alleles. All of the lines tested contained the *MIPS1a* allele (Table 3). We also tested two lines representing divergent germplasm: 'Peking' and PI88788 (Bilyeu and Beuselinck 2005). Both of these lines also contained the *MIPS1a* allele. Further testing of other soybean lines led to the identification of several lines that contained the *MIPS1b* allele (Table 3).

Expression profile of four soybean *MIPS* genes

Similar to the genomic sequence data for the *MIPS* gene family, only a limited amount of data was available in the literature for the expression of *MIPS* genes in soybean. Previous work via Northern blot analysis concluded that *MIPS1* was highly expressed in seeds, but it was not examined if *MIPS2*, *MIPS3* or *MIPS4* were also expressed in seeds (Hegeman et al. 2001). It is essential to know which genes in the *MIPS* gene family are expressed in seeds since these genes will be targeted for mutation screening in order to obtain a low phytate phenotype in seeds. As a first step in determining the expression pattern of the four *MIPS* genes, an 'electronic Northern' was assembled based on EST frequency in the GenBank database. As shown in Table 4, *MIPS1* ESTs were isolated from a variety of tissues indicating that *MIPS1* is expressed in more than one tissue. However, the relative frequency of *MIPS1* ESTs suggests that *MIPS1* expression is highest in cotyledons. These data are consistent with previous results that *MIPS1* is highly expressed in developing seeds (Hegeman et al. 2001). The isolation of *MIPS2* ESTs from a variety of tissues indicates that *MIPS2* is also expressed in more than one tissue. The relative frequency of these ESTs might suggest that *MIPS2* is expressed highest in roots. *MIPS3* and *MIPS4* are also expressed in several tissues with their highest EST frequency from leaves and seedlings. Taken together, these data suggest that *MIPS* genes are expressed in a variety of tissues with *MIPS1* being the predominant gene in developing seeds.

Table 4: Classification of *Glycine max MIPS* ESTs

	No. EST sequences encoding			
Library source	*Gm*MIPSl	*Gm*MIPS2	*Gm*MIPS3	*Gm*MIPS4
Vegetative				
Roots	4	8	0	0
Leaves	1	5	25	10
Seedlings	0	2	39	24
Other	2	3	7	10
Reproductive				
Flowers/meristems	1	3	14	7
Seed coats	4	6	2	0
Cotyledons	15	4	2	0
Hypocotyl and plumules	0	2	0	0
Pods	2	1	1	3
Somatic embryos	1	0	0	0

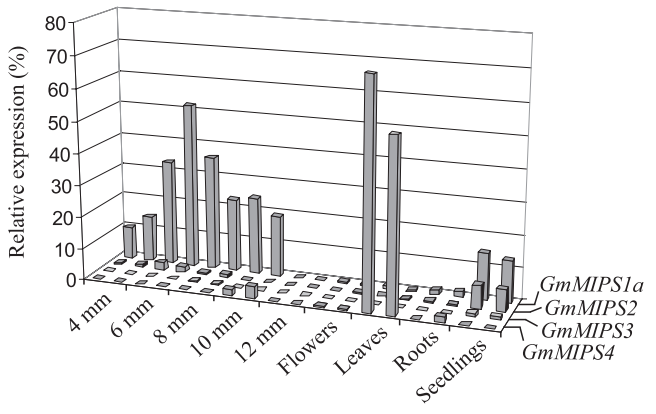

Fig. 3: Relative expression of soybean *MIPS* genes in staged flowers, developing seeds, and vegetative tissues. Steady state 'Williams 82' mRNA levels for each gene were quantitated by real-time PCR following reverse transcription of total RNA from each tissue. The two histograms for each sample represent independent RNA isolations and the percent of each gene normalized to the housekeeping gene, *PEPC16*. For developing seed tissues, the length of the seed is given as an indication of developmental stage

To experimentally examine *MIPS* expression in soybean tissues, quantitative real-time RT-PCR was conducted on all four *MIPS* genes (Fig. 3). The previously characterized house keeping gene *PEPC16* was used as an internal standard (Tuteja et al. 2004). *MIPS1* exhibited its highest expression in developing seeds with expression peaking around 6 mm seeds. While *MIPS1* was either not expressed or minimally expressed in 12 mm seeds, flowers, leaves and roots, *MIPS1* was moderately expressed in seedlings. *MIPS2* and *MIPS3* were poorly expressed in all tissues examined with a slightly higher expression in seedlings. *MIPS4* was poorly expressed in all tissues examined except in leaves where it was highly expressed. These data indicate that *MIPS1* is highly expressed and *MIPS2*, *MIPS3* and *MIPS4* are poorly expressed in developing seeds.

Characterization of the *MIPS* gene family in a low phytate soybean

A publicly available low phytate soybean line has been developed as part of a mutagenesis breeding program: CX1834 (which includes the genetically related lines CX1834-1-2, CX1834-1-3 and CX1834-1-6, all selected for their low phytate phenotype). The CX1834 lines are derivatives of the mutant M153 and have a 50% reduction in seed phytate levels with a concomitant increase in available phosphorus (Oltmans et al. 2003, Hulke et al. 2004, Walker et al. 2006). The low phytate trait in this line has been demonstrated genetically to be the result of two independent loci (Oltmans et al. 2003, Walker et al. 2006). Because CX1834 shows no accumulation of phytic acid precursors, a phenotype common to *MIPS* mutants, a candidate gene approach with the *MIPS* genes was pursued (Raboy et al. 2000, Pilu et al. 2003, Shukla et al. 2004). All four of the *MIPS* cDNAs were amplified, cloned, and sequenced from CX1834. No mutations were identified in any of the *MIPS* sequences, even in the seed-expressed *MIPS1* gene. The *MIPS1a* allele was found to be present in CX1834 when genomic DNA was queried. These data indicate that the CX1834 low phytate phenotype is not due to coding mutations in any of the four *MIPS* genes.

The maize lpa1 (low phytic acid 1) mutation has been mapped to a *MIPS* gene, although no sequence mutations are present in the *MIPS* coding sequence (Shukla et al. 2004). Thus, a non-coding mutation, such as a promoter or enhancer mutation, is most likely the cause of the low phytic acid phenotype. To investigate whether a non-coding mutation in *MIPS1* is the cause of the low phytic acid phenotype in CX1834, we investigated the association of the CX1834 *MIPS1a* allele with the low phytate trait in progeny of a cross between CX1834 and '5601T' (Pantalone et al. 2003), a line containing normal phytate levels and the *MIPS1b* allele. A population was developed for which available phosphorus was quantified and scored for each of the progeny lines. Utilizing the polymorphism in *MIPS1*, we queried the lines for the presence of the *MIPS1a* allele from CX1834 or the *MIPS1b* allele from '5601T' and compared it with the available phosphorus phenotype. No association was found with the *MIPS1a* allele from CX1834 and the available phosphorus phenotype. Since two QTL markers (Walker et al. 2006) have been developed for the low phytic acid phenotype of CX1834, these markers were also followed in the progeny of the cross. However, no correlation between the *MIPS1* genotype and either of the two markers was found (data not shown). Taken together, these data indicate that a *MIPS1* related mutation, coding or non-coding (promoter, enhancer, splice-site mutation, etc), is not involved in the CX1834 low phytate phenotype.

Discussion

In this work, we carried out a comprehensive characterization of the soybean *MIPS* gene family. Ultimately, our results will facilitate the use of reverse genetics and candidate gene approaches to investigate pathways involved in *myo*-inositol signalling and phytic acid biosynthesis. One goal of this work was to clear up the confusion surrounding the *MIPS1* gene, which has been deposited in GenBank as two different sequences differing by six nucleotides (Hegeman et al. 2001, Hitz et al. 2002). Our results with 'Williams 82' were consistent with the *MIPS1* sequence reported by Hitz et al. 2002. It is not clear why the *MIPS1* sequence from Hegeman et al. 2001, is different from the sequence reported here and by Hitz et al. 2002, but there is no evidence in the soybean EST collection for the amino acid differences reported in the original Hegeman sequence.

In addition to characterizing the coding sequence of *MIPS1*, the complete genomic sequence for all four *MIPS* genes was determined. All of the soybean *MIPS* genes share a common intron/exon structure. Thus, the soybean *MIPS* genes are highly conserved at both the sequence and structural level. This was not surprising as all *MIPS* genes to date are highly conserved genes (Majumder et al. 1997, 2003, Bachhawat and Mande 2000). The high sequence conservation among the genes should be given consideration when it is desirable to distinguish among the genes.

We also set out to address the identity of the *MIPS1b* allele, which is found in a subset of soybean cultivars. Previous work indicated that besides a deletion of 39 bp in the final intron, *MIPS1b* contained 40 bp changes in the coding sequence (representing seven amino acid differences) to *MIPS1a* (Hitz et al. 2002). However, when all of the *MIPS* genes from a cultivar containing the *MIPS1b* allele were investigated, no differences in the coding sequence between *MIPS1a* and

MIPS1b were found. The non-coding sequence differences comprised a 39-bp deletion in the final intron and a single base pair deletion in the sixth intron. One potential reason why Hitz et al. 2002 reported differences in the coding sequence between *MIPS1a* and *MIPS1b*, is that Hitz et al. 2002 inadvertently isolated and sequenced *MIPS2* cDNA, which contains 42 base changes (representing seven amino acid differences) to *MIPS1* for 'Williams 82'. As Hitz et al. 2002 reported 40 base changes (representing seven amino acid differences) it is possible that a specific allele of *MIPS2* was sequenced, as the particular cultivar that was used in this analysis was not listed.

The expression profiles of the *MIPS* genes were also determined. The results indicate that individual members of the soybean *MIPS* gene family are temporally and spatially expressed. Through quantitative real-time RT-PCR analysis, it could be determined that only *MIPS1* is highly expressed in developing seeds, as was previously suggested by Northern blot analysis (Hegeman et al. 2001). Interestingly, Hegeman et al. 2001, found *MIPS1* expression to peak around 2–4 mm stage whereas peak expression around 6–8 mm stage was found. It is not clear why there is a discrepancy with regards to *MIPS1* peak expression. The expression analysis also revealed that despite their sequence conservation to *MIPS1*, *MIPS2* and *MIPS3* are poorly expressed in a variety of soybean tissues, and *MIPS4* is highly expressed only in leaves. These data indicate that while there is a great deal of conservation of the gene itself, the regulatory elements of the gene have diverged.

Even with an extensive EST dataset available for soybean, the contrast between experimental expression analysis and 'electronic Northern' analysis is substantial. While 'electronic Northern' analysis correctly predicted high *MIPS4* expression in leaves, it did not indicate that *MIPS2* and *MIPS3* are poorly expressed in a variety of tissues. Thus, while 'electronic northern' analysis can be informative at the preliminary level, actual experimental analysis must be completed to properly determine expression profiles.

Development of the genomic resources for the soybean *MIPS* genes allowed to investigate them as candidate genes for mutations in the low phytate line CX1834. Soybean line CX1834 was reported to have a phenotype similar to the previously described soybean *MIPS* mutant (Hitz et al. 2002, Oltmans et al. 2003, Hulke et al. 2004, Walker et al. 2006). All four *MIPS* genes from CX1834 were sequenced but no mutations were identified in the coding sequences.

To detect whether a non-coding mutation in *MIPS1* is the cause of the low phytic acid phenotype in CX1834, the association of the CX1834 *MIPS1a* allele with the low phytate phenotype and the associated QTL markers (Walker et al. 2006) in descendants of a cross between CX1834 and a line that contained normal phytate levels and the *MIPS1b* allele were investigated. No association was found with the *MIPS1a* allele from CX1834 and the available phosphorus phenotype or either of the two QTLs. In conclusion *MIPS1* is not responsible for the low phytate phenotype in line CX1834.

To summarize, the *MIPS* soybean gene family is characterized by sequencing of all four genes and studying their expression profile in a variety of tissues. These data will aid in any reverse genetic strategies used to alter *MIPS* gene expression. Given that *MIPS1* is highly expressed in developing seeds, this gene should be targeted for mutagenesis to identify mutants that contain low seed phytic acid. To this end, TILLING is currently used (McCallum et al. 2000a,b, Colbert et al. 2001) to identify mutations in *MIPS1* with the ultimate goal of characterizing low phytic acid soybean lines.

References

Bachhawat, N., and S. C. Mande, 2000: Complex evolution of the inositol-1-phosphate synthase gene among archaea and eubacteria. Trends Genet. **16,** 111—113.

Barillot, E., B. Lacroix, and D. Cohen, 1991: Theoretical analysis of library screening using a N-dimensional pooling strategy. Nucleic Acids Res. **19,** 6241—6247.

Bilyeu, K. D., and P. R. Beuselinck, 2005: Genetic divergence between North American ancestral soybean lines and introductions with resistance to soybean cyst nematode revealed by chloroplast haplotype. J. Hered. **96,** 593—599.

Bilyeu, K. D., L. Palavalli, D. Sleper, and P. Beuselinck, 2005: Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. Crop Sci. **45,** 1830—1836.

Bouis, H. E., 2000: Special issue on improving human nutrition through agriculture. Food Nutr. Bull. **21,** 351—576.

Brinch-Pedersen, H., L. D. Sorensen, and P. B. Holm, 2002: Engineering crop plants: getting a handle on phosphate. Trends Plant Sci. **7,** 118—125.

Bruno, W. J., E. Knill, D. J. Balding, D. C. Bruce, N. A. Doggett, W. W. Sawhill, R. L. Stallings, C. C. Whittaker, and D. C. Torney, 1995: Efficient pooling designs for library screening. Genomics **26,** 21—30.

Colbert, T., B. J. Till, R. Tompa, S. Reynolds, M. N. Steine, A. T. Yeung, C. M. McCallum, L. Comai, and S. Henikoff, 2001: High-throughput screening for induced point mutations. Plant. Physiol. **126,** 480—484.

Cromwell, G. L., and R. D. Coffey, 1991: Phosphorus-a key essential nutrient, yet a possible major pollutant – its central role in animal nutrition. In: T. P. Lyons (ed.), Biotechnology in the Feed Industry, 133—145. Alltech Technical Publishers, Nicholasville, Kentucky.

Gizlice, Z., T. E. Carter, and J. W. Burton, 1994: Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci. **34,** 1143—1151.

Hegeman, C. E., L. L. Good, and E. A. Grabau, 2001: Expression of D-myo-inositol-3-phosphate synthase in soybean. Implications for phytic acid biosynthesis. Plant. Physiol. **125,** 1941—1948.

Hitz, W. D., T. J. Carlson, P. S. Kerr, and S. A. Sebastian, 2002: Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant. Physiol. **128,** 650—660.

Hulke, B. S., W. R. Fehr, and G. A. Welke, 2004: Agronomic and seed characteristics of soybean with reduced phytate and palmitate. Crop Sci. **44,** 2027—2031.

Irvine, R. F., and M. J. Schell, 2001: Back in the water: the return of the inositol phosphates. Nat. Rev. Mol. Cell. Biol. **2,** 327—338.

Johnson, M. D., 1994: The *Arabidopsis thaliana myo*-inositol 1-phosphate synthase (EC 5.5.1.4). Plant. Physiol. **105,** 1023—1024.

Johnson, M. D., and D. Burk, 1995: Isozyme of 1L-myo-inositol-phosphate synthase from Arabidopsis (accession no. U30250; PGR95-067). Plant. Physiol. **109,** 721.

Johnson, M. D., and I. M. Sussex, 1995: 1 L-*myo*-inositol 1-phosphate synthase from *Arabidopsis thaliana*. Plant. Physiol. **107,** 613—619.

Klein, P. E., R. R. Klein, S. W. Cartinhour, P. E. Ulanch, J. Dong, J. A. Obert, D. T. Morishige, S. D. Schlueter, K. L. Childs, M. Ale, and J. E. Mullet, 2000: A high-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. Genome Res. **10,** 789—807.

Lackey, K. H., P. M. Pope, and M. D. Johnson, 2003: Expression of 1L-myoinositol-1-phosphate synthase in organelles. Plant. Physiol. **132,** 2240—2247.

Loewus, F. A., and M. W. Loewus, 1983: *myo*-inositol: its biosynthesis and metabolism. Annu. Rev. Plant Physiol. **34,** 137—161.

Loewus, F. A., and P. P. N. Murthy, 2000: *myo*-inositol metabolism in plants (review). Plant Sci. **150,** 1—19.

Lott, J. N. A., 1984: Accumulation of seed reserves of phosphorus and other minerals. In: D. R. Murray (ed.), Seed Physiology, Vol. 1, 140—166. Academic Press, Australia.

Majumder, A. L., M. D. Johnson, and S. A. Henry, 1997: 1L-myo-inositol-1-phosphate synthase. Biochim. Biophys. Acta. **1348,** 245—256.

Majumder, A. L., A. Chatterjee, K. Ghosh Dastidar, and M. Majee, 2003: Diversification and evolution of L-myo-inositol 1-phosphate synthase. FEBS. Lett. **553,** 3—10.

McCallum, C. M., L. Comai, E. A. Greene, and S. Henikoff, 2000a: Targeting induced local lesions in genomes (TILLING) for plant functional genomics. Plant. Physiol. **123,** 439—442.

McCallum, C. M., L. Comai, E. A. Greene, and S. Henikoff, 2000b: Targeted screening for induced mutations. Nat. Biotechnol. **18,** 455—457.

Oltmans, S. E., W. R. Fehr, G. A. Welke, and S. R. Cianzio, 2003: Inheritance of low-phytate phosphorus in soybean. Crop Sci. **44,** 433—435.

Pantalone, V. R., F. L. Allen, and D. Landau-Ellis, 2003: Registration of '5601T' soybean. Crop Sci. **43,** 1123—1124.

Pilu, R., D. Panzeri, G. Gavazzi, S. K. Rasmussen, G. Consonni, and E. Nielsen, 2003: Phenotypic, genetic and molecular characterization of a maize low phytic acid mutant (lpa241). Theor. Appl. Genet. **107,** 980—987.

Prattley, C. A., and D. W. Stanley, 1982: Protein-phytate interactions in soybeans: I. Localization of phytate in protein bodies and globoids. J. Food Biochem. **6,** 243—253.

Raboy, V., 1997: Accumulation and storage of phosphate and minerals. In: B. A. Larkins, and I. K. Vasil (eds), Cellular and Molecular Biology of Plant Seed Development, 441—477. Kluwer Academic, Dordrecht, Netherlands.

Raboy, V., 2002: Progress in breeding low phytate crops. J. Nutr. **132,** 503S—505S.

Raboy, V., D. B. Dickinson, and F. E. Below, 1984: Variation in seed total P, phytic acid, zinc, calcium, magnesium, and protein among lines of *Glycine max* and *G. sojae*. Crop Sci. **24,** 431—434.

Raboy, V., P. F. Gerbasi, K. A. Young, S. D. Stoneberg, S. G. Pickett, A. T. Bauman, P. P. N. Murthy, W. F. Sheridan, and D. S. Ertl, 2000: Origin and seed phenotype of maize low phytic acid 1–1 and low phytic acid 2–1. Plant. Physiol. **124,** 355—368.

Shukla, S., T. T. VanToai, and R. C. Pratt, 2004: Expression and nucleotide sequence of an INS (3) P-1 synthase gene associated with low-phytate kernels in maize (*Zea mays* L.). J. Agric. Food Chem. **52,** 4565—4570.

St. Martin S. K., A. J. Calip-DuBois, R. J. Fioritto, A. F. Schmitthenner, D. B. Min, T.-S. Yang, Y. M. Yu, R. L. Cooper, and R. J. Martin, 1996: Registration of 'Ohio FG1' soybean. Crop Sci. **36,** 813.

Strother, S., 1980: Homeostasis in germinating seeds. Ann. Bot. **45,** 217—218.

Tuteja, J. H., S. J. Clough, W. C. Chan, and L. O. Vodkin, 2004: Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in *Glycine max*. Plant Cell **16,** 819—835.

Walker, D. R., A. M. Scaboo, V. R. Pantalone, J. R. Wilcox, and H. R. Boerma, 2006: Genetic mapping of loci associated with seed phytic acid content in CX1834-1-2 soybean. Crop Sci. **46,** 390—397.

Wilcox, J. R., G. S. Premachandra, K. A. Young, and V. Raboy, 2000: Isolation of high seed inorganic P, low-phytate soybean mutants. Crop Sci. **40,** 1601—1605.

Zhou, J. R., and J. W. Jr Erdman, 1995: Phytic acid in health and disease. Crit. Rev. Food. Sci. Nutr. **35,** 495—508.